Mathematical Application to the Analysis of Indoor Air

Witzar Destin
Faculty Mentor: Dr. Umesh Nagarkatte
Department of Physical, Environmental and Computer Sciences
Medgar Evers College, City University of New York

-

ABSTRACT

Chemometrics is the use of mathematical, statistical or graphical methods to analyze and measure the chemical composition of a substance. Related measurements of the composition are taken, and values of interest are inferred through these methods (Lavine, 1998, p. 209). In these experiments, curve fitting is being studied in order to develop a repeated pattern in the graphs of indoor air samples. Air samples vary from area to area because of different chemicals existing in the air. The samples are collected and then analyzed by gas chromatography using a Gas Chromatography Mass Spectrometer (GC/MS). Numerical data is collected in terms of peak numbers, retention time, area, start and end times of the peaks. Each chemical compound in the sample elutes at a definite time, called the retention time, as the sample passes through the chromatographic column gives rise to a peak, and area under which is computed by the GCMS software.

From the data, peak areas of the blank sample are subtracted to eliminate the effect of the sample medium, and since Toluene, a gasoline component, is the most predominant and frequently occurring chemical compound in these samples, the area of Toluene divides all other areas. The relative areas, most of which are now less than one, are plotted against the retention times. Curve fitting is performed and polynomials are fitted to the graphs and parts of the graphs. The coefficients of the polynomials are then compared.

PROJECT DESCRIPTION

The goal is to apply mathematics to find a fingerprint for indoor air. The steps to attain the goal are as follows.

1	Collecting air samples
2	Chromatographic processing
3	Data collection
4	Preliminary Data Analysis using Excel
5	Curve fitting using RS/1

Another team of research assistants using Tetraethylene glycol dimethyl ether (Tetraglyme) to scrub the Volatile Organic Compounds (VOC's) from the air conducted air sampling. Laboratory analysis of the air samples using purge and trap and GC/MS yields chromatograms. Figure 1a.is a chromatogram of one air sample. It shows the abundance versus the retention time of the sample indicated by the peaks. The chromatograms are then examined, graphs are created and curves fitted for analysis of the data.

MATERIALS AND METHODS

The materials used for collecting air sample were SKC air sampling pump, rubber tubes, flow meter, impinger, glass tubes, thermos, ice and rock salt. Instruments used to identify the Volatile Organic Compounds (VOC's), trapped in the medium, Tetraethylene glycol dimethyl ether (Tetraglyme), are the Hewlett-Packard Purge and Trap Concentrator and Hewlett-Packard AP-6890 Series Mass Selective Detector.

Tetraglyme is used as a scrubber to remove VOC's from the air. Aliquots of the Tetraglyme, 100 micrograms, are dissolved in 5 ml of water. The solution is injected into the Purge and Trap Concentrator, Helium or other inert gas is used to purge and trap the air sample where it is heated and the pumped into the Gas Chromatography/Mass Spectrometer (GCMS) and a chromatogram is produced.

The materials used for mathematical analysis of the chromatogram are Microsoft Excel and RS/1 software, books and articles on Statistics, Chemometrics and Gas Chromatography were studied extensively.

Data collected from the chromatogram of various indoor air samples in given area of a building at given time consists of peak #, retention time, area, start and finish time (or width). The data is extracted on a disk as a text file and then processed as an excel file. Table 1 is the unprocessed data that is extracted from the chromatogram. The data is then exported into the RS/1 software for in-depth analysis and curve fitting.

DISCUSSION

The first step in the analysis of the data is a visual observation of the chromatograms of the different samples from various locations for any recognizable repeating pattern that will help in the determination of the curve that would the best fit for the data. All retention time greater than 17.00 minutes were ignored, as it was determined that those peaks could be from the medium and not the sample. A database of 18 gas chromatograms of air samples from four different locations was used in the study.

The data is then processed, using MS Excel, which involves considering columns of peak #, retention time and area only, by removing columns of data not required for present investigation. Using the RS/1, a graph is created from the numerical table of the gas chromatogram and a visual comparison is done; three data samples were used for the initial analysis. Visual observation of the graphs indicates that in the approximate range between 5 and 13 minutes a second-degree polynomial would be the most likely fit.

The analysis of the next group of data involves the use of a "blank" (which is the medium without any air sample). Figure 1b is a chromatogram of a blank. Its peaks are similar to that of the sample, but with less abundance. This is the basis for calculating the area difference of the graphs. Toluene consistently elutes at the same time up to two decimal places Its area was therefore used to calculate the relative area; all other areas of the samples were divided by the area of Toluene (see Table 2).

The chromatograms and tables of the raw data shows that there are three clusters of activity with respect to retention time. As a result of this, three graphs were created for every sample data—one graph for every cluster of activity.

The first graph was created using the data points in the retention time range from 3.117 to 9.693 minutes (see Graph 1). Another graph was created for the range that is greater than or equals to 11.0, but less than or equals to 12.3 (see Graph 2). The third graph's range is between 12.389 and 14.5 (see Graph 3).

The method of least squares was used to fit polynomials of different degrees and select the polynomials that are visibly close to the distribution of the data set. There are three polynomials for each graph—polynomials of third to fifth-degrees. A second-degree polynomial is used for the entire graph (see Graph 4). All the samples were treated as such, and the fifth-degree polynomial provided the best fit to the data points of the graphs.

To further analyze Gas Chromatographic data, since curve fitting to the data has been established, the next step in the research should in the direction of the other areas of Chemometrics. Areas such as pattern recognition and principal component analysis, statistical discriminant analysis and the Mahalanobis distance which other researchers have used to find a fingerprint and the source of chemical compounds in data samples.

According to Lavine (1995), "Principal component analysis (PCA) is a method for transforming the original measurement variables into new, uncorrelated variables called principal components" (p. 3848). He further stated that "Often, the two or three largest principal components of the data will capture the bulk of the variance or information, hence, we can use them to generate a plot that represents the structure of the *p*-dimensional measurement space" (p. 3848). Another researcher, De Maesschalck (2000) that " In the original variable space, the MD takes into account the correlation in the data, since it is calculated using the inverse of the variance-covariance matrix of the data set of interest" (p. 2)

These two papers have shown how successful PCA can be.

CONCLUSION

1. The set of polynomials fitting the graphs of air samples from a particular area has their coefficients close at least up to the hundredth place.
2. Air samples from different locations produce different sets of polynomials.
3. The analysis produces a mathematical model of indoor air.
4. Since there are other chemical compounds within the areas of the peaks, further research needs to be conducted to determine what other chemical compounds are present. The use of new methods, procedures, software and the studying of accepted and published literature in the field of Chemometrics may be helpful in this regard.

REFERENCES

De Maesschalck, R., Jouan-Rimbaud, D. & Massart, D. L. (2000). The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems 50 1-18

Lavine, B. K. (1995). Source identification of underground fuel spills by pattern recognition analysis of high-speed gas chromatograms. Analytical Chemistry, 67, (21) 3846-3852.

Lavine, B. K. (1998). Chemometrics. Analytical Chemistry, 70, (12) 209R-228R.